Information needs and knowledge acquisition in context of online disease-related information search

He Fang
PhD candidate of Department of Information Management, Peking University, China
Exchange Student of Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore
heather_f@163.com

Supervisors: Wang yanfei
Department of Information Management, Peking University, China
yfwang@pku.edu.cn

Christopher Khoo
Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore
chriskhoo@pmail.ntu.edu.sg

ABSTRACT
Objectives: The objectives of this study are: (1) to identify how situations affect prior knowledge and information needs; (2) to identify how prior knowledge and information needs affect information behavior.
Methods: 2 focus groups and 20 pilot interviews were conducted firstly. The interview outline was derived from focus groups (hold in July 12, 13) and literature, and the whole interview included three sessions: the pre-search session, the search session and the post-search. Then a questionnaire survey (N=377) and individual interviews (80) will be conducted.
Keywords: Information needs, prior knowledge, disease-related information search, knowledge acquisition

INTRODUCTION
A large and growing number of people use Internet to find disease-related information, such as symptoms, treatment, and prevention (Elizabeth, Pam, Peter & Lesley, 2007; Fox, 2011). According to Accenture Consulting’s 2016 survey about Chinese people, more than half of the participants prefer to get the disease related information through the Internet. The Internet is becoming the main source for people to obtain disease-related information and they often select search engines as their channel preference (De Choudhury, Morris, 2014). According to Sense-making Theory, information search is a knowledge acquisition process that involves sense making and bridging of the knowledge gap (Dervin, 1992). The knowledge gap, namely information needs arise from uncertainty when users encounter problems or barriers. The way user defines knowledge gap is typically associated with situation and the user’s existing medical knowledge. Although some research on information behaviour of cancer patients and their family members have paid attention to the disease situation, few studies
have been conducted on general illnesses which are more common, and few have focused on how situation, info needs and prior knowledge interacts.

LITERATURE REVIEW

Situations and Information Needs
Some previous studies had attempted to investigate information needs to build a universal(trans-situational) model to predict information behaviour, then with a manifestation of a recognition that trans-situational prediction has not been productive, Paisley, Allen (1968) and Dervin(1986) called for a situational orientation. Situation is a combination of various factors, including search for whom, what problems users have, etc., which constitutes the search context and leads to search behaviour. Sense-making theory place the situation in a significant position, providing a user concerned perspective rather than information resource construction from information system perspective. In the latest version of the sense-making model which labelled “Situation-Gap-Use” (Dervin, 2001), information seeker stopped when facing barriers implied in a situation, and the gap between the aims user want to realized and the existing knowledge emerged. According to this model, the sense-maker firstly generalize gap in mind, and then try to built a bridge across the gap. “Information need” links to the gap essentially, as it has been defined similarly: 1) when a person recognizes something wrong in his or her state of knowledge and wishes to resolve the anomaly (Belkin, 1978); 2) when internal sense runs out (Dervin, 1977; 1980).

However, information need is a very general concept that often be viewed differently according to research perspectives. It can be summarized in several categories and also can be fully explored at the cognitive level. According to Dervin, the situational essence of the information needs lie in how situations are assumed to constrain further movements of users which can be coded. She primarily proposed categories which described the nature of the movement stop as decision (facing a road with two or more branches ahead), problematic (being dragged down a road not of user’s own choosing), or spin-out (having no road). Meanwhile, Hall and Dervin suggested that information search was the interaction between internal cognition and external behaviour, and individuals shaped their cognitive maps accordingly through the seeking process, which means sense making process was dynamic and depends on case by case knowledge structure and new information that users got.

When study involved with large-sample investigation, as knowledge that can meet information needs can’t be specifically described before information seeking, the information need is actually an aim of seeking bridge across a gap. To some extent information needs is misunderstood and misused, one solution for this is when study the relationship of situations and information needs, we can firstly generalize “information needs” as aims to verify the correlation between situational factors and information needs (aims), then focus on concrete knowledge acquisition through search to meet their information needs.

Prior knowledge and information retrieval
Prior knowledge is derived from the field of Learning Theory, refers to existing knowledge towards a phenomenon or a topic. It is suggested that people generally reflection influenced by not only the environment but also from existing knowledge and expectations (Remez et al. Rubin et, 1981; al., 1997). Research shows that the existing knowledge has effects on information retrieval in many aspects, including the impact on the length of user search terms (Hsieh-Yee, 1993; Wildemuth, 2004), the effect of retrieval results reading time (Kelly, 2006a, 2006b; Kelly & Cool, 2002), and the influence on the precision index and recall index of search performance (Allen, 1991; Marchionini, 1989). Klein (1998) believed that user
experience was also a kind of information source, through the identification of new models of situations to fill information gaps, if this failed, users turned to seek information externally. The methods design of above research typically had two approaches, one was the comparison of experience or domain knowledge level of different groups, and the groups were divided according to areas of expertise, scores of knowledge tests or conducted self-ratings; the other is to compare the same group of users’ search behaviour under different period, such as before and after a course of study or training. Hsieh-Yee (1993) gave participants a familiar field of search tasks and tasks outside of one of these areas (low familiarity). She found that when users faced with an unfamiliar topic, they relied more on thesauri while chose search terms. Hembrooke, Granka, Gay, and Liddy (2005) took a similar method, asking the experts to select two familiar tasks and two unfamiliar tasks, much longer and more complex terms were used when they searched for familiar ones. Duggan and Payne (2008) of participants, divided participants into the high and low groups according to the knowledge test scores, and they found that the music knowledge had little effect on the search, but the football knowledge had a great influence on the user search, and the performance, accuracy is positively associated with the prior football knowledge. Measuring and variables setting is the main challenge for existing knowledge research, generally there are four methods used in these study: (1) self evaluates the overall level of prior knowledge by users; (2) tests the familiarity of users by thesaurus; (3) tests users by questionnaires about a domain’s knowledge; (4) analyse the summary or report wrote by users at the end of search. Regard to the reliability of knowledge self assessment, Cole (2010) had one group of participants self-assessed their knowledge on genetics tasks, and then used MeSH vocabulary to judge their knowledge, he found the two results had a significant positive correlation. Kelly, Kantor, Morse, Scholtz and Sun (2006) asked experts to self evaluate and mutual evaluate, and they suggested experts who wrote high quality reports were very accurate when evaluating other expert knowledge, but tended to underestimate their own knowledge. In spite of the bias of self evaluation, it is still a relatively reliable method to know the level of users’ prior knowledge. The above research was mostly carried out in the context of academic information seeking, in terms of search for disease, Kirkwood (1999) mentioned, for user who got well understanding of disease history, if the situation was similar to the situation in the past, the user's actions and decisions would depend on the experience of the past. However, the measurement method of disease related knowledge, and whether it has similar influences on the information needs like academic information seeking tasks have received little concerns.

**METHODOLOGY DESIGN**

This study developed a coding scheme by grounded theory and partly from previous research. The following sections give detailed information on participants, the research procedure, data collection and analysis tools, and the coding scheme.

**Participants**

Participants of this study need meet the following qualifications:

- aged from 18 to 40 years;
- with an education level of undergraduate students or a Bachelor’s degree;
- no medical education background;
- no mental disease.
The education criterions were used to guarantee participants’ Internet skill and access of Internet is similar, and they are layperson without medical education background, because professionals are suggested to have very different information needs. The last criterion ensured that participants did not have mental diseases, as this has not been fully studied, search for mental illnesses is more likely to differ with physical illnesses.

**Procedure, Data Collection and Analysis Tools**

2 focus groups and 20 pilot interviews were conducted firstly. The semi-structured interview outline was derived from 2 focus group (hold in July 12, 13) and literature. Then based on transcripts of focus groups and pilot interviews, a scheme was developed by Grounded Theory coding mechanism and partly from the literature. A questionnaire survey was conducted. The questionnaire was designed on the basis of coding scheme and was sent to about 377 participants by email or SNS (Wechat) to figure out the main situational factors which influence prior knowledge information needs, so far, we got 128 respondents.

According to the information needs, select the individual interviews participants, and the interview included three sessions: the pre-search session, the search session and the post-search. In the search session, participants were asked to conduct a search within 25 mins, and the processes were captured and audiotaped by screen recording software (Morae).

**The Derivation of Coding Scheme**

The coding scheme (see table 1) used in this study was divided into three categories:

- variables of situation;
- variables of prior knowledge;
- information needs variables.
The factors of situation includes history of disease, seriousness of disease, the state of disease and search for whom. The most common factors are history of disease, seriousness of disease. It’s also suggests that the state of disease is an important factor. Chronic diseases, surgery, trauma and cancer are studied respectively by Chinese medical practitioners in the hospital. After comparing and analyzing 27 Chinese papers, some commonalities were found, although the disease is different, the newly diagnosed patient and their family members want to understand the diagnosis, and patients prepared to surgical operation need treatment risk information, besides this, their family members also want to know how take care of the patient. So I add “perceived state of disease” and “search for whom” into coding scheme. The state of disease refers to user perceived state such as undiagnosed, diagnosed and being treated (medical treatment), diagnosed and daily management (diet and exercise). The relationship between information searchers and patients, which often be overlooked, has been proved to influence reading contents and total retrieval time (Wenjing Pian & Christopher Khoo, 2016), and it can reasonably supposed that the closeness is associated significantly with prior knowledge from the pilot interviews. In terms of prior knowledge, it was divided into two parts, one is knowledge of patients and the other is general knowledge of disease, as mentioned in literature review, self-rating is a verified and often adopted method with high credibility. In pilot interviews, the existing knowledge often be described specifically as following: symptoms, etiology, types of disease, prognosis, treatment, so these are coded to the scheme, but these code are only used in interviews.

<table>
<thead>
<tr>
<th>Variable group</th>
<th>Variable name</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>situations</td>
<td>search for whom</td>
<td>1. for oneself 2. for family members and close friends 3. for ordinary friends</td>
</tr>
<tr>
<td></td>
<td>perceived state of disease</td>
<td>1. symptoms known but not diagnosis 2. diagnosis known but not treatment 3. being treated 4. daily management</td>
</tr>
<tr>
<td></td>
<td>seriousness</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>chronic or not</td>
<td>yes, no</td>
</tr>
<tr>
<td>prior knowledge</td>
<td>overall knowledge about patient</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>overall knowledge about disease</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>etiology</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>types of disease</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>prognosis</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td></td>
<td>treatment</td>
<td>Likert scale 1-7</td>
</tr>
<tr>
<td>information needs</td>
<td>information needs (aims)</td>
<td>self-diagnose, etiology, comparison of treatment options, making prognosis, disease monitoring and management, etc.</td>
</tr>
<tr>
<td></td>
<td>the urgency of info need</td>
<td>Likert scale 1-7</td>
</tr>
</tbody>
</table>
Information needs have some common features since the uncertainty of situation and health decisions which individuals need to eliminate and make are similar in many ways. This is supported by the data from pre-search interview and search session, which indicated that information needs could be generalized into systematic aims, such as self-diagnose, comparison of treatment options, making prognosis, disease monitoring and management, etc.

**Hypothesis**

So far, the hypotheses are as following:

1. search for whom and the state of disease correlate with overall knowledge of patient (correlation analysis);
2. search for whom and illness seriousness correlate with the urgency of need (correlation analysis);
3. search for whom and the state of disease are the main factors that influence information.

In terms of knowledge acquired in context of online disease-related information search, we have not set any hypothesis, expect to find in the the interview/search.

**RESULTS (NOT COMPLETE)**

**Demography**

The summary of demographic information is given in Table 2. All participants are undergraduate or above, ages of 68% participants are between age 20-30.

<table>
<thead>
<tr>
<th>Education background</th>
<th>Undergraduate: 42</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Master degree: 41</td>
</tr>
<tr>
<td></td>
<td>PhD: 45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female:78</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male:50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>18-20 yrs: 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-30 yrs: 87</td>
</tr>
</tbody>
</table>

**Survey Results (not complete)**

Group by search for whom, see as Table 3, 38.3% of the participants search for their own health issue, with an average seriousness of disease 3.7 and prior knowledge of 3.9, and 49% users of this group have not been diagnosed which is the highest among three groups.

48.4% of the participants search for their family members or closed relatives and friends, in only 6.45% of the cases, the diseases have not been diagnosed, and their average seriousness of disease is 4.8 and prior knowledge is 4.1, both are higher than those who search for themselves.

13.3% of the participants search for ordinary friends and relatives, in this group all disease has been diagnosed and with the highest seriousness scores of 6.7, conversely, the prior knowledge score is the lowest of 2.6.
Table 3. General table of situation, overall prior knowledge and information needs (N=128)

<table>
<thead>
<tr>
<th>search whom</th>
<th>perceived state of disease</th>
<th>seriousness(average)</th>
<th>prior knowledge overall rating</th>
<th>aims</th>
<th>the urgency of info need</th>
</tr>
</thead>
<tbody>
<tr>
<td>for oneself (n=49)</td>
<td>1. symptoms known but not diagnosis(n=24); 2. diagnosis known but not treatment(n=11); 3. being treated(n=14); 4. daily management(n=0)</td>
<td>3.7 (1-7)</td>
<td>3.9 (1-7)</td>
<td>self-diagnose, (49%) etiology (18.3%), comparison of treatment options(6.1%), making prognosis (26.7%)</td>
<td>5.25 (1-7)</td>
</tr>
<tr>
<td>for family members and close friends (n=62)</td>
<td>1. symptoms known but not diagnosis(n=4); 2. diagnosis known but not treatment(n=20); 3. being treated(n=37); 4. daily management(n=1)</td>
<td>4.8 (1-7)</td>
<td>4.1 (1-7)</td>
<td>self-diagnose (6.5%), etiology (22.6%), comparison of treatment options (6.5%), making prognosis (51.4%), disease monitoring and management (13%)</td>
<td>5.6 (1-7)</td>
</tr>
<tr>
<td>for ordinary friends (n=17)</td>
<td>1. symptoms known but not diagnosis(n=0); 2. diagnosis known but not treatment(n=1); 3. being treated(n=16); 4. daily management(n=0)</td>
<td>6.7 (1-7)</td>
<td>2.6 (1-7)</td>
<td>etiology (23.6%), making prognosis (76.4%)</td>
<td>5 (1-7)</td>
</tr>
</tbody>
</table>

REFERENCES


